

DW/DM Using Microsoft SQL Server

2012

(CSCI 4144/6405)

TA: Mayank Malhotra (malhotra@cs.dal.ca)

January 2014

Outline

- A quick re-cap of DW/DM concepts
- Set-up of SQL Server 2012
- Build a data cube
- OLAP operators for DW in SQL Server
- DM model in SQL Server
 - Building a clustering model
 - Building a decision tree mining model for classification

A quick re-cap of DW/DM concepts

- Multi-dimensional data model (cube) and DW schema
- OLAP operations
- Clustering
- Classification

A quick re-cap of DW/DM concepts

- Multi-dimensional data model
 - Data cube
- DW schema
 - Star
 - A fact table + a few dimension tables (one for each dimension)
 - Snowflake
 - A fact table + a few dimension tables (one or more for dimension)
 - Reduce redundancies by normalization

A quick re-cap of DW/DM concepts

- Measure
 - Attributes in Fact table;
 - Can be aggregated (e.g sale, production, population)
- Dimension
 - Attributes based on which measures can be aggregated (e.g. time, product, country)

A quick re-cap of DW/DM concepts

- OLAP operations
 - Roll-up (Generalization)
 - Drill-down (Specification)
 - Slice (data selection based on predicate of one dimension)
 - Dice (slice on two or more dimensions)

A quick re-cap of DW/DM concepts

- Clustering

- Cluster or group data without class-label based on some principles (e.g maximizing intra-class similarity, minimizing the inter-class similarity)

- Classification

- Use class-labelled data to construct a model to predict the class of data whose class label is unknown;
- Decision Tree

Outline

- A quick re-cap of DW/DM concepts
- Set-up of SQL Server 2012

- Build a data cube
- OLAP operators for DW in SQL Server
- DM model in SQL Server
 - Building a decision tree mining model for classification
 - Building a clustering model

SQL Server 2012 is available:

- Installed on machines in Lab1 / Lab2 CS
-
- HelpDesk (borrow an installation disk)
 - Download from <https://msdn.cs.dal.ca/>
 - Installation Guide (Disclaimer alert):
<http://www.slideshare.net/samehd/sql-server-2012-install>

Set-up of SQL Server 2012

- **Login (Windows based authentication)**
 - Username: **CS_Account**
 - Password: **XXXXXX**

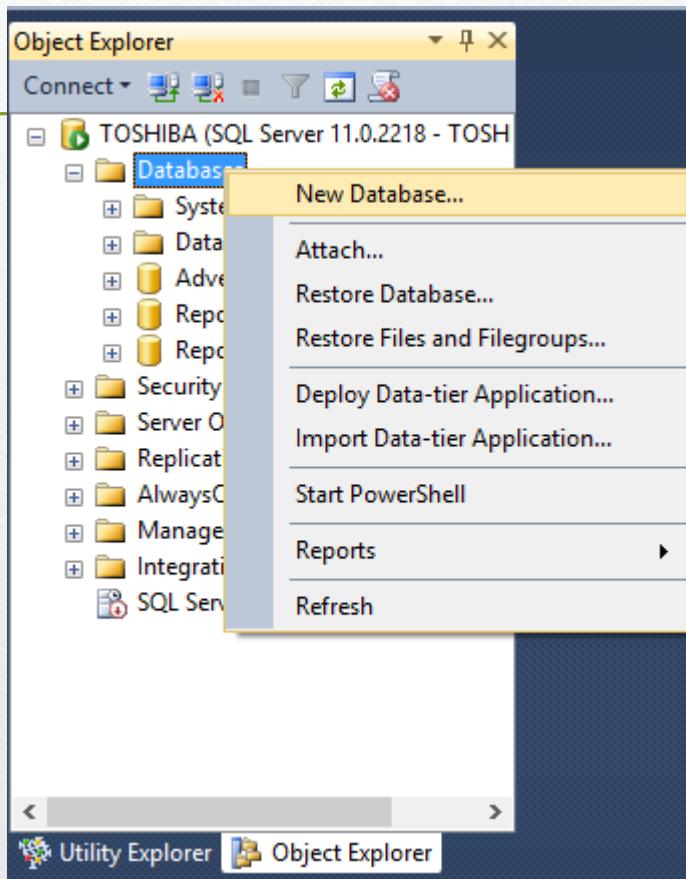
Set-up of SQL Server 2012

Attach Database to SQL server 2012

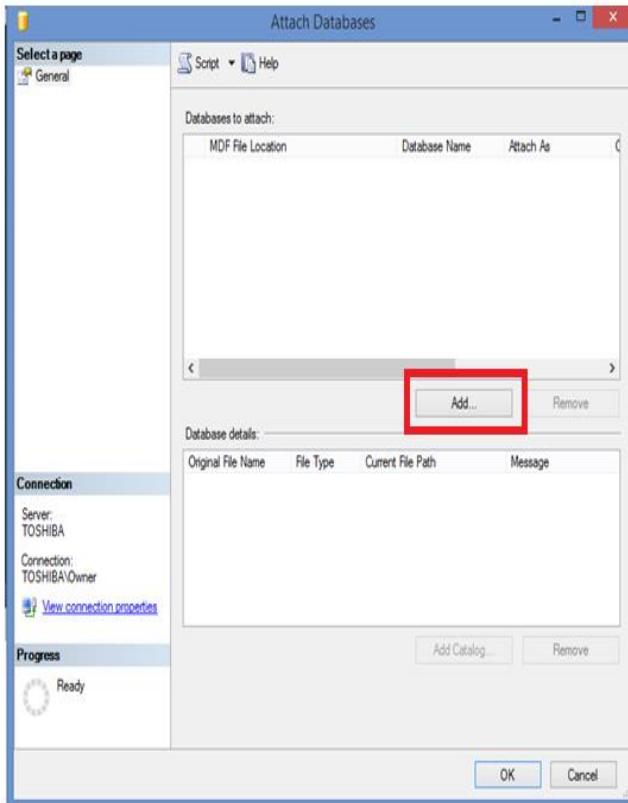
You can download “AdventureWorks2012” database from
<http://msftdbprodsamples.codeplex.com/releases/view/93587>

- Copy .mdf file of your database.
- Go to All programs-> Program Files->Microsoft SQL Server->MSSQL11.MSSQLSERVER->MSSQL->DATA
- Paste your .mdf file

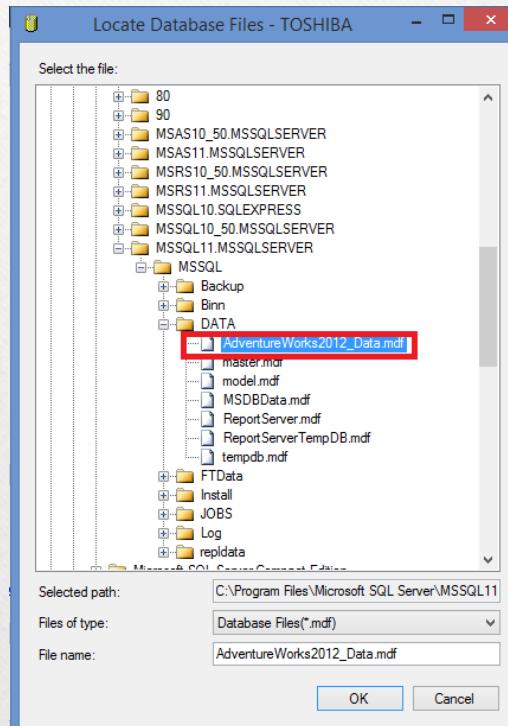
- Now Open SQL Server Management Studio
- Right click on Database and click on “attach”.



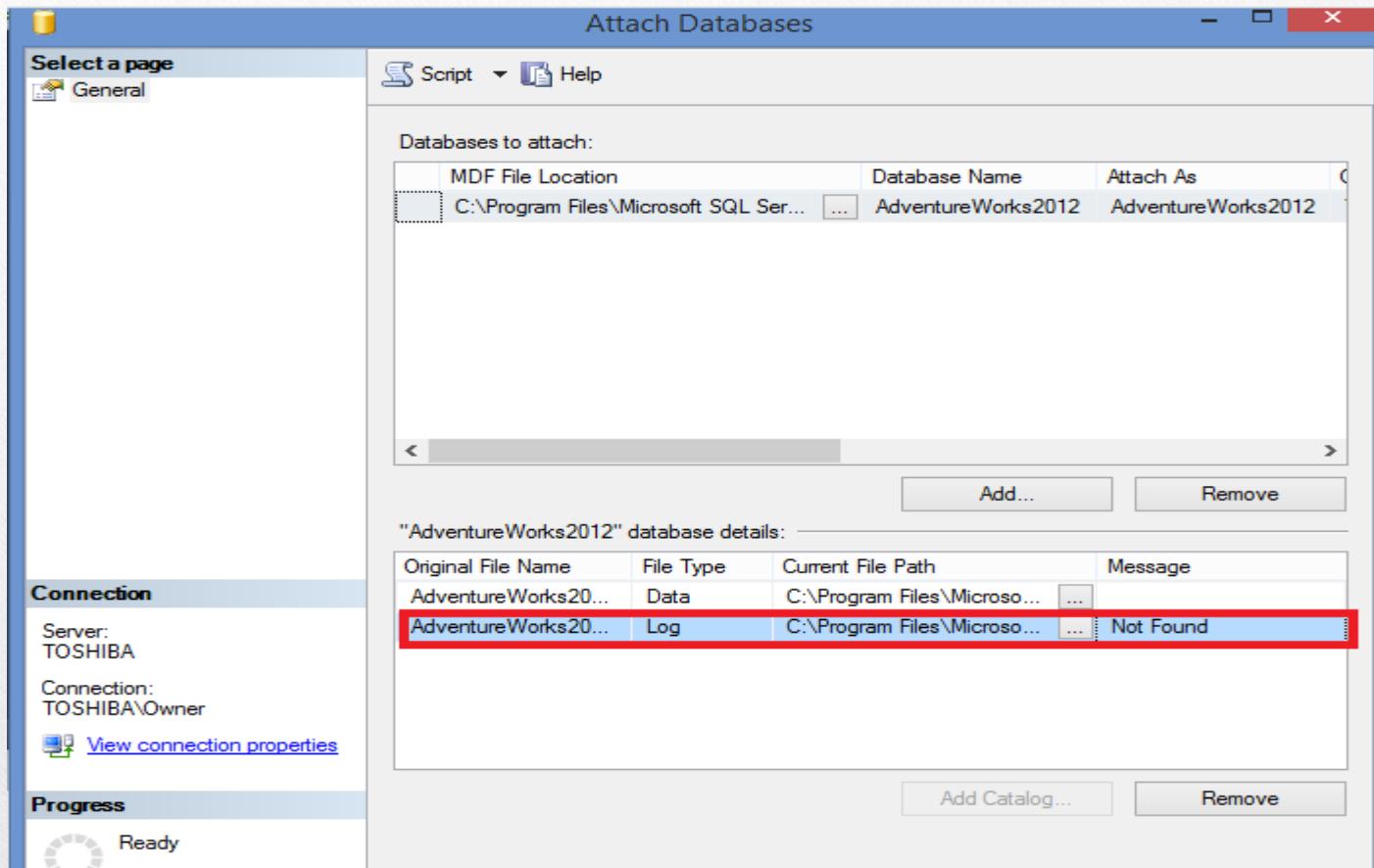
- Click on “ADD”.



- Browse : Program Files->Microsoft SQL Server->MSSQL11.MSSQLSERVER->MSSQL->DATA
- select the Database from the given list and press “ok”.



- select “log file” and click on “remove” and then click “ok”



Open Business Intelligence Development Studio

*Start menu → All Programs --> Microsoft SQL Server
2012 -->SQL server data tools-> SQL Server Business
Intelligence Development Studio*

Set-up of SQL Server 2012

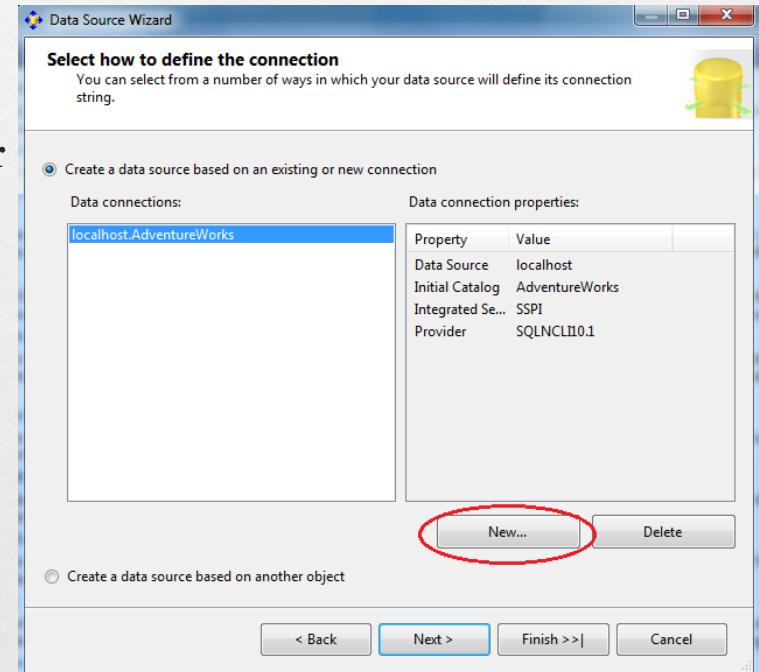
- Create an Analysis Service Project
 - File → New → Project

 - Choose *Analysis Service Project* template
 - Select the **Location** where you can keep it for future access (saving on your local drive is highly recommended)
 - (i.e. memory stick, U: \\)
- Open an existing Analysis Service Project
 - File → Open → Project/Solution
 - Select the project you stored before

Set-up of SQL Server 2012

- ***Data Source Wizard***

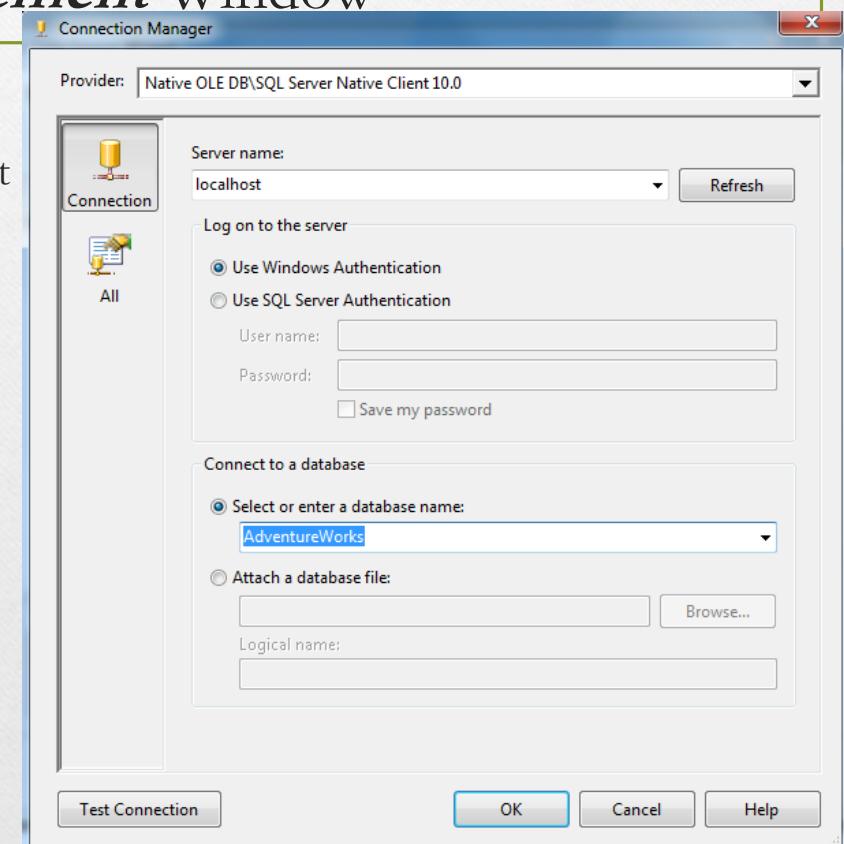
- Right click Data Sources → New Data Sources...
- Create a data source based on an existing or new connecting
- Click **New** button
 - go to the **Connection Manager**



Set-up of Microsoft SQL Server

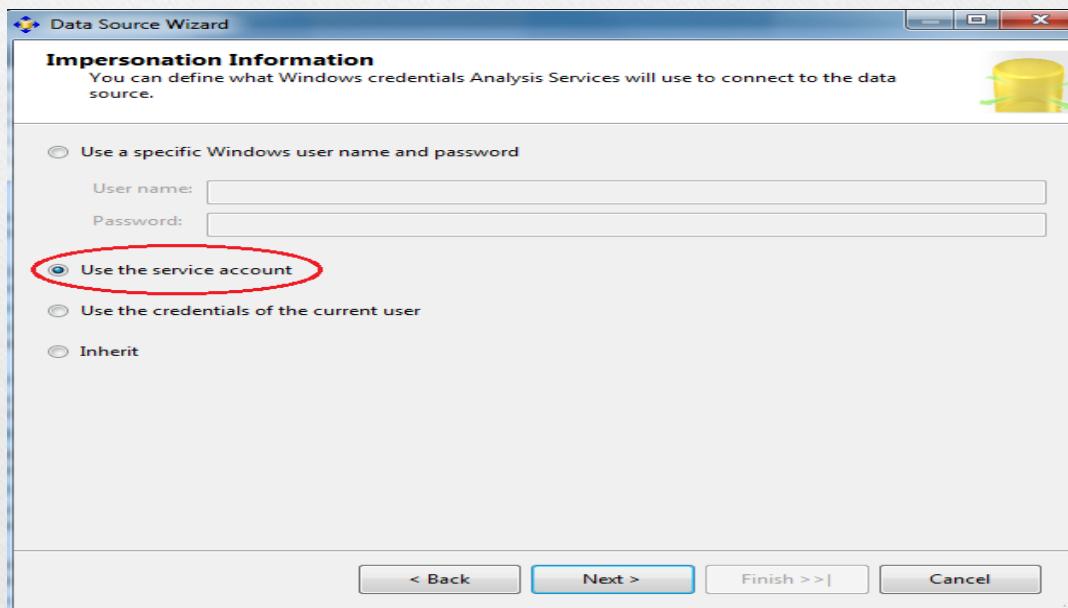
Data Source Wizard(cont.)

- In the *Connection Management* Window
 - Select server
 - ie. MACHINENAME /localhost
 - Click **Test Connection**
 - Select database
 - example database
 - ie. Adventureworks



Set-up of Microsoft SQL Server

- Data Source Wizard(cont.)
 - In the following **Impersonation Information** dialog, choose “*use the service account*”.
 - Select “Next” until reaching the end of wizard.



Set-up of Microsoft SQL Server

- **Create Data Sources Viewer**
 - Right click **Data Sources Views** in Solution Explorer, open **Data Sources Views Wizard**.
 - In **Selection Tables and Views**, you can select the tables to be viewed. You can use >> to select all the tables in the current databases for viewing.
 - **Note:** Even though the assignment does not require you to create data source viewer, I strongly recommended you to do so before OLAP and data mining since it provides a convenient reference to relationships between tables in the databases that may be useful in data warehousing and data mining.

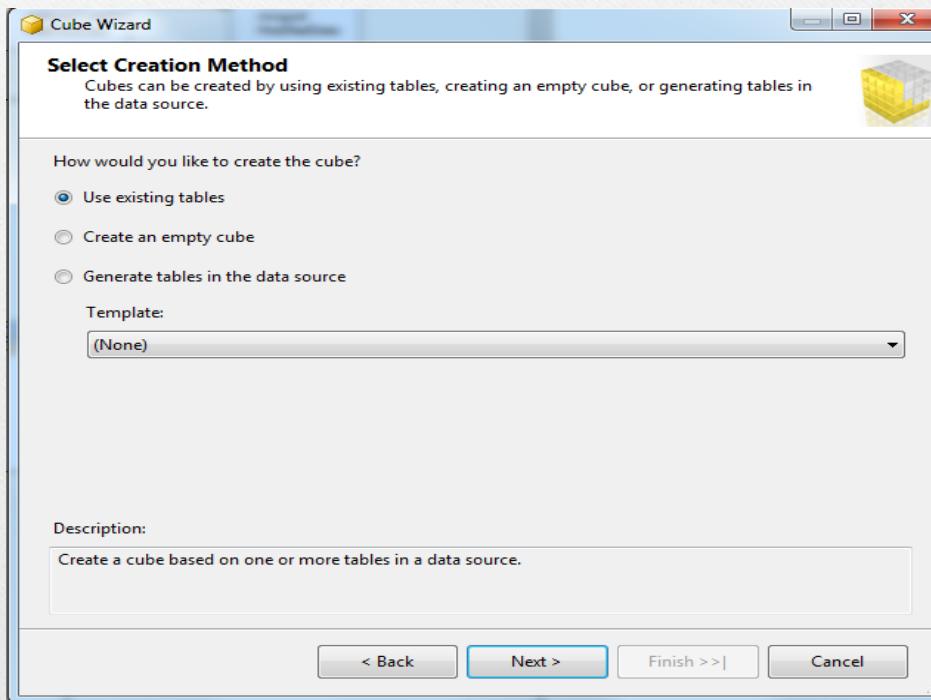
Outline

- A quick re-cap of DW/DM concepts
- Set-up of SQL Server 2012
- Build a data cube
- OLAP operators for DW in SQL Server
- DM model in SQL Server
 - Building a decision tree mining model for classification
 - Building a clustering model

Build a data cube

- **Create Data Cube Objects**

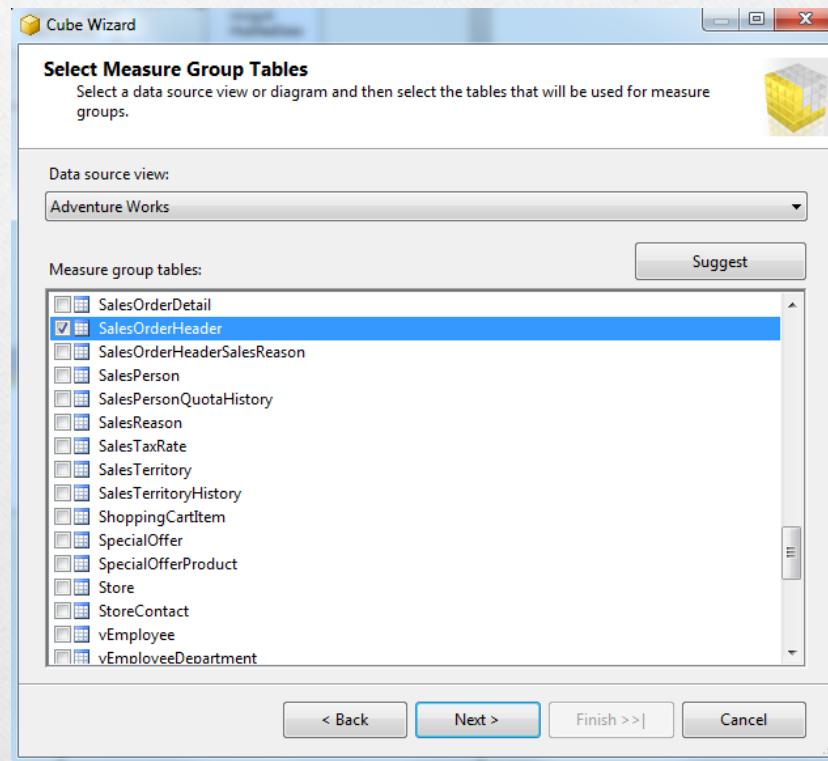
- Right click **Cube** folder and click “New Cube” in order to enter the **Cube Wizard**.
- In **Select Creation Method**, choose “*Use existing tables*”,



Build a data cube

Cube Wizard

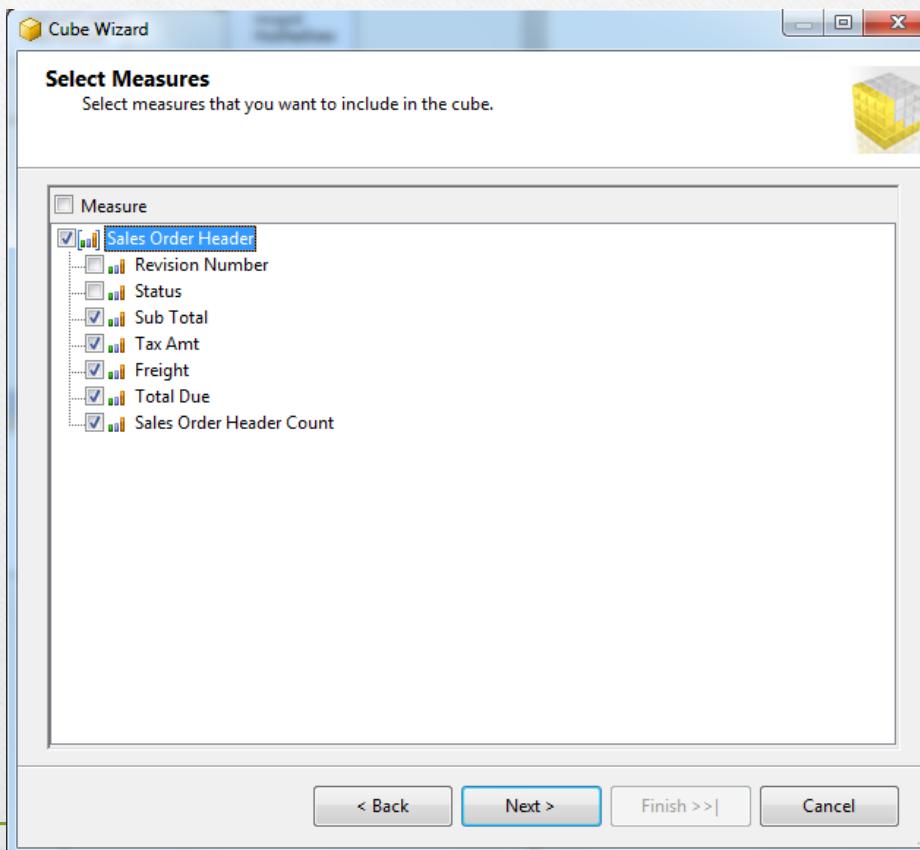
- In Data source view, choose “Adventure Works”;
- In Measure group tables, tick “SalesOrderHeader”
 - As the fact table



Build a data cube

Cube Wizard (cont.)

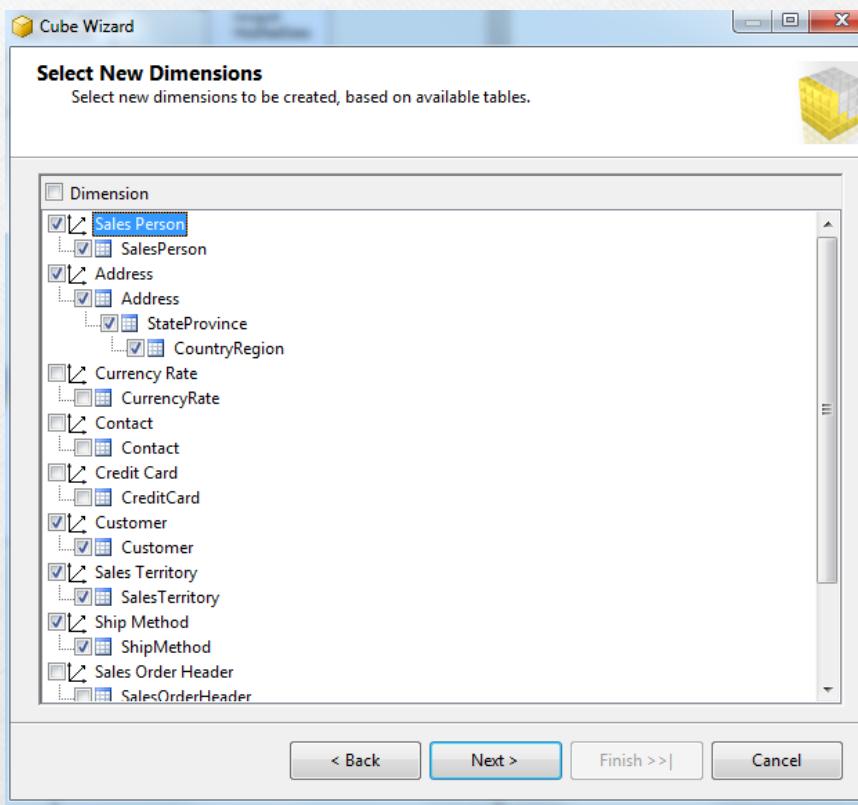
- In **Select Measures**, choose the measures



Build a data cube

Cube Wizard (cont.)

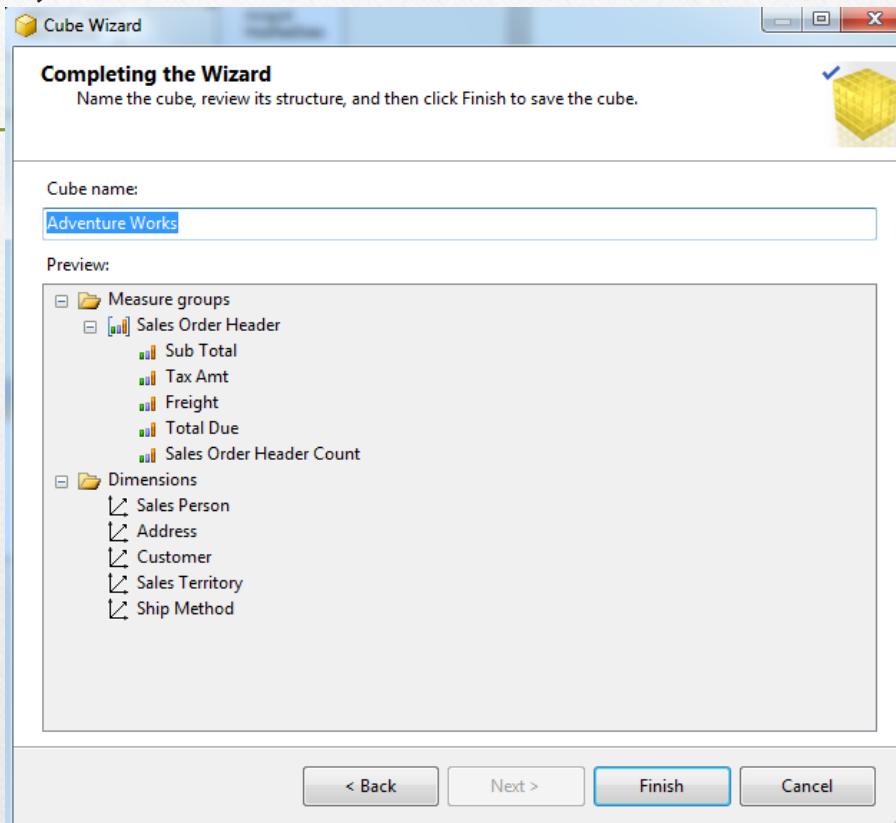
- In **Select New Dimensions**, choose the dimensions



Build a data cube

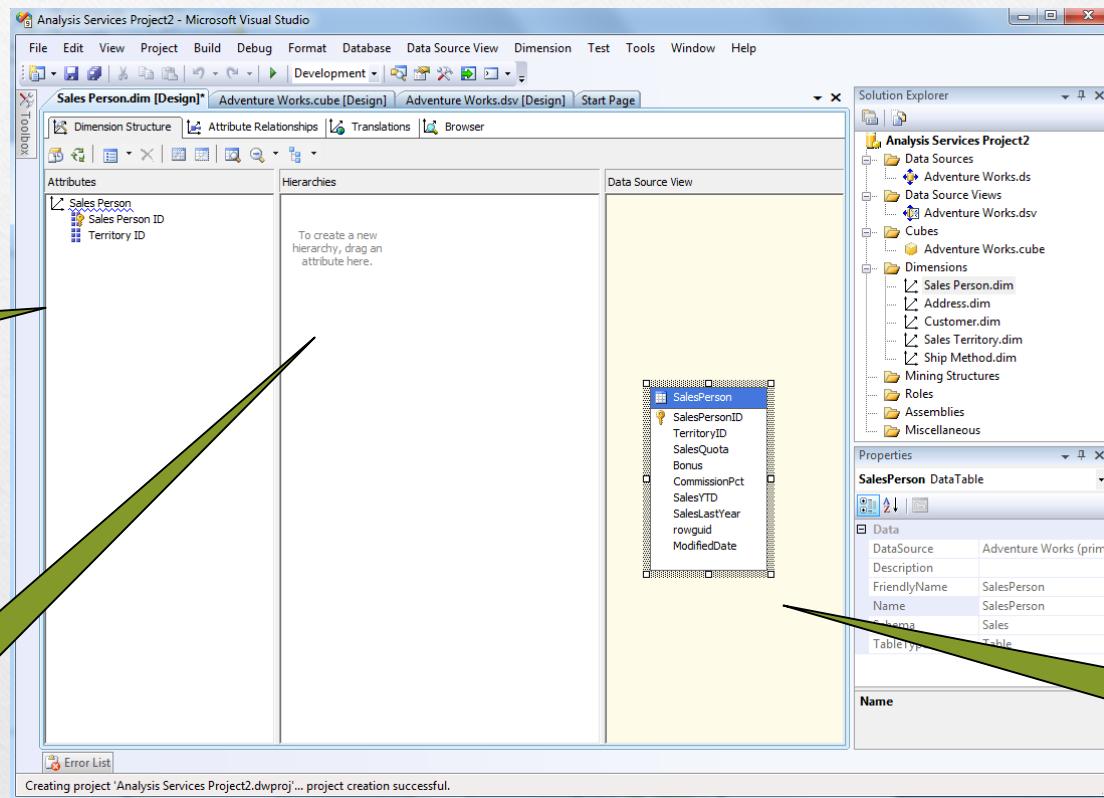
- **Cube Wizard (cont.)**

- Enter Cube name
- Finish



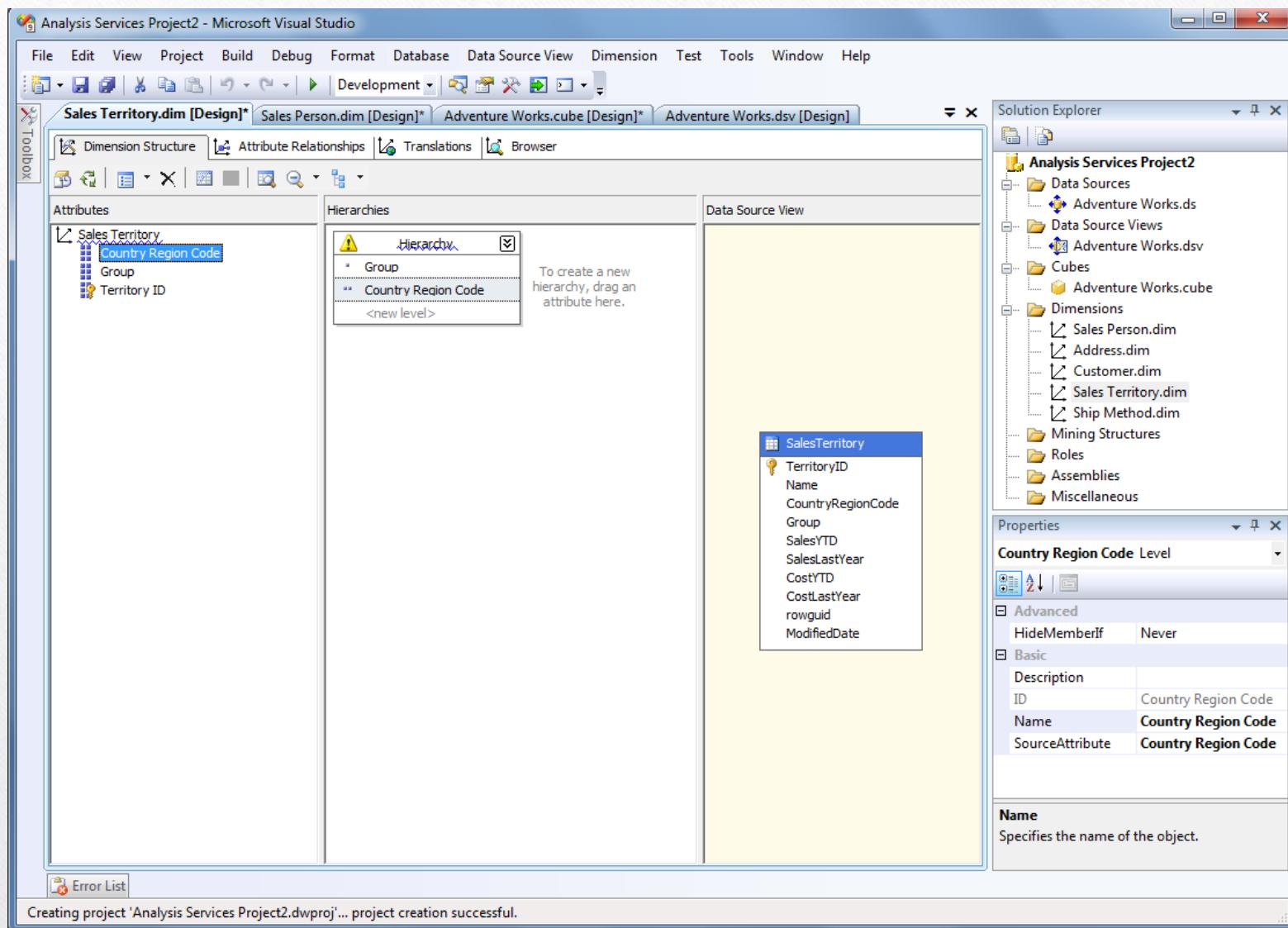
Build a data cube

- Dimension modification
 - Open Dimension Designer
 - double-click a dimension name in the Dimensions node of Solution Explorer.



Build a data cube

- Adding Attributes to the *Sale Territory* dimension
 - Double-click the dimension name
 - Drag the following columns from the table in the Data Source View pane to the Attributes pane:
 - **Group**
 - **CountryRegionCode**
- Creating a Hierarchy on a dimension (*Sale Territory*)
 - Drag the *Group* attribute from the Attributes pane into the Hierarchies pane.
 - Drag the *Country Region Code* attribute from the Attributes pane into the <new level> cell in the Hierarchies pane, underneath the *Group* level.



Build a data cube

- **Create Data Cube Objects (cont.)**

- Before viewing the cube, you need to process both the Cubes object and Dimensions object. Right-click the Cube you have created and choose “**Process**” and start to “**run**” the processing of the cube object. Similarly, right-click Sale Territory dimension and choose “**Process**” and start to “**run**” the processing of the dimension object.
- After the processing of both cube and dimension objects, right-click the name of your cube object, say it is called “Adventure Works” and select “**Browse**”.

Outline

- A quick re-cap of DW/DM concepts
- Set-up of Microsoft SQL Server 2012
- Build a data cube
- OLAP operators for DW
- DM model in SQL Server
 - Building a clustering model
 - Building a decision tree mining model for classification

OLAP operators for DW

- In left-most tree-like list, you can select the measures from the fact table for viewing. Right-click the measures you want to view and select “Add to data area”. The selected measures will be shown in the middle panel.
- In the middle panel, you can specify the expression(s) for dimension attributes in order to view the measures that satisfy the given expression(s).
 - For example, you can give the expression like “Customer Territory Equal 10” to view, for instance, the Total Sale order Count in that region.
 - In this way, you can easily perform the basic OLAP operations such as drill down, roll up, slice and dice for data exploration from the cube.

Analysis Services Project2 - Microsoft Visual Studio

File Edit View Project Build Debug Database Cube Test Tools Window Help

Customer.dim [Design] Sales Territory.dim [Design] Sales Person.dim [Design] Adventure Works.cube [Design]

Perspective: Adventure Works Language: Default

Measure Group: <All>

Adventure Works

- Measures
 - Sales Order Header
 - Freight
 - Sales Order Header Count
 - Sub Total
 - Tax Amt
 - Total Due
 - Bill To Address
 - Bill To Address - Territory
 - Customer
 - Customer - Territory
 - Customer - Territory.Country Region C
 - Customer - Territory.Group
 - Customer - Territory.Territory ID**
 - Customer - Territory.Hierarchy
 - Sales Person
 - Sales Person - Territory
 - Sales Territory
 - Ship Method
 - Ship To Address
 - Ship To Address - Territory

Dimension Hierarchy Operator Filter Expression

Dimension	Hierarchy	Operator	Filter Expression
Customer - Territory	Customer - Territory.Territory ID	Equal	{ 10 }

Drop Filter Fields Here

Drop Column Fields Here

Sales Order Header Count
188

Drop Row Fields Here

Error List

Ready

Solution Explorer

- Analysis Services Project2
 - Data Sources
 - Adventure Works.ds
 - Data Source Views
 - Adventure Works.dsv
 - Cubes
 - Adventure Works.cube
 - Dimensions
 - Sales Person.dim
 - Address.dim
 - Customer.dim
 - Sales Territory.dim
 - Ship Method.dim
 - Mining Structures
 - Roles
 - Assemblies
 - Miscellaneous

Deployment Progress - Analysis Ser... Server: localhost Database : Analysis Services Project2

Command

Status:

Deployment Completed Successfully

Properties Deployment Progress

OLAP operators for DW

- Roll-up
 - Total Due of all Customer Territory group

Dimension	Hierarchy	Operator	Filter Expression
Sales Territory	Sales Territory.Hierarchy	Equal	{ All }
<Select dimension>			
<input type="button" value="Drop Filter Fields Here"/>			
<input type="button" value="Drop Column Fields Here"/>			
Total Due			
ds Here	108266245.7018		

OLAP operators for DW

- Drill-down
 - “Total Due of in different Customer Territory groups”

Drop Filter Fields Here						
	Group ▾	Country Region Code ▾				
Fields Here	<input checked="" type="checkbox"/> Europe	<input type="checkbox"/> North America	CA	US	Total	<input checked="" type="checkbox"/> Pacific
	Total Due	Total Due			Total Due	<input checked="" type="checkbox"/> Grand Total
	14715218.6417	19316293.7919	72049623.6031	91365917.3949999	2185109.6651	108266245.7018

OLAP operators for DW

- Slice
 - “Total Due of ***Customer Territory*** Equals Europe & North America”
- Dice
 - “Total Due of ***Customer Territory*** Equals Europe & North America and ***Sales Person ID*** begins with 28”

Dimension	Hierarchy	Operator	Filter Expression
Sales Territory	Sales Territory.Hierarchy	Equal	{ Europe, North America }
<Select dimension>			
Drop Filter Fields Here			
Group	Country Region Code		
Europe	North America	Grand Total	
Total Due	Total Due	Total Due	
14715218.6417	91365917.3949999	106081136.0367	

Dimension	Hierarchy	Operator	Filter Expression
Sales Territory	Sales Territory.Hierarchy	Equal	{ Europe, North Amer }
Sales Person	Sales Person ID	Begins With	28
<Select dimension>			
Drop Filter Fields Here			
Group	Country Region Code		
Europe	North America	Grand Total	
Sales Person ID	Total Due	Total Due	Total Due
280		4453081.8838	4453081.8838
281		8761727.2886	8761727.2886
282	5169355.2923	2798413.5083	7967768.8006
283		5029846.9145	5029846.9145
284	985641.9261		985641.9261
285		11384512.9896	11384512.9896
286	6083690.9556		6083690.9556
287		3110529.5797	3110529.5797
289	2476530.4677		2476530.4677
Grand Total	14715218.6417	35538112.1645	50253330.8062

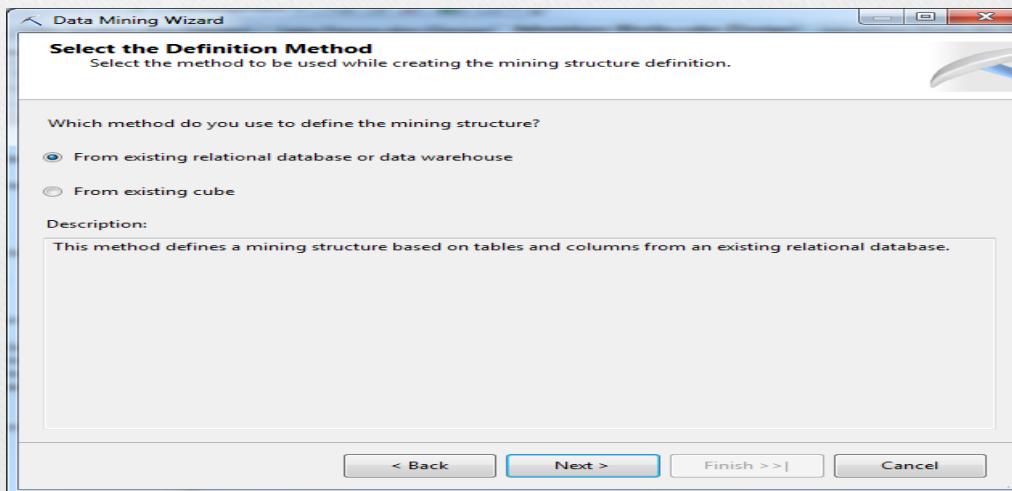
Outline

- A quick re-cap of DW/DM concepts
- Set-up of Microsoft SQL Server
- Build a data cube
- OLAP operators for DW
- DM model in SQL Server
 - Building a clustering model
 - Building a decision tree mining model for classification

DM models in SQL Server

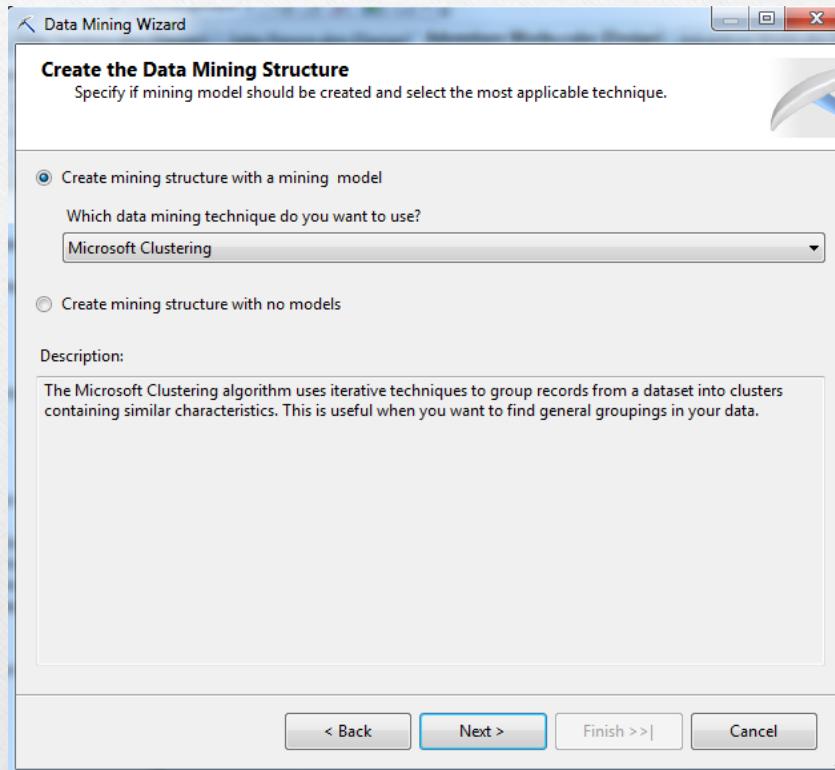
- **Create Data Mining Objects**

- Use **Data Mining Wizard**
 - Right-click the **Mining Structure Folder** in the Solution Explorer and then choose **New Mining Structure**.
 - First, choose whether you are creating a model from a relational or multidimensional source (an OLAP cube).
 - In our case, select "**From existing relational database or data warehouse**".



DM models in SQL Server

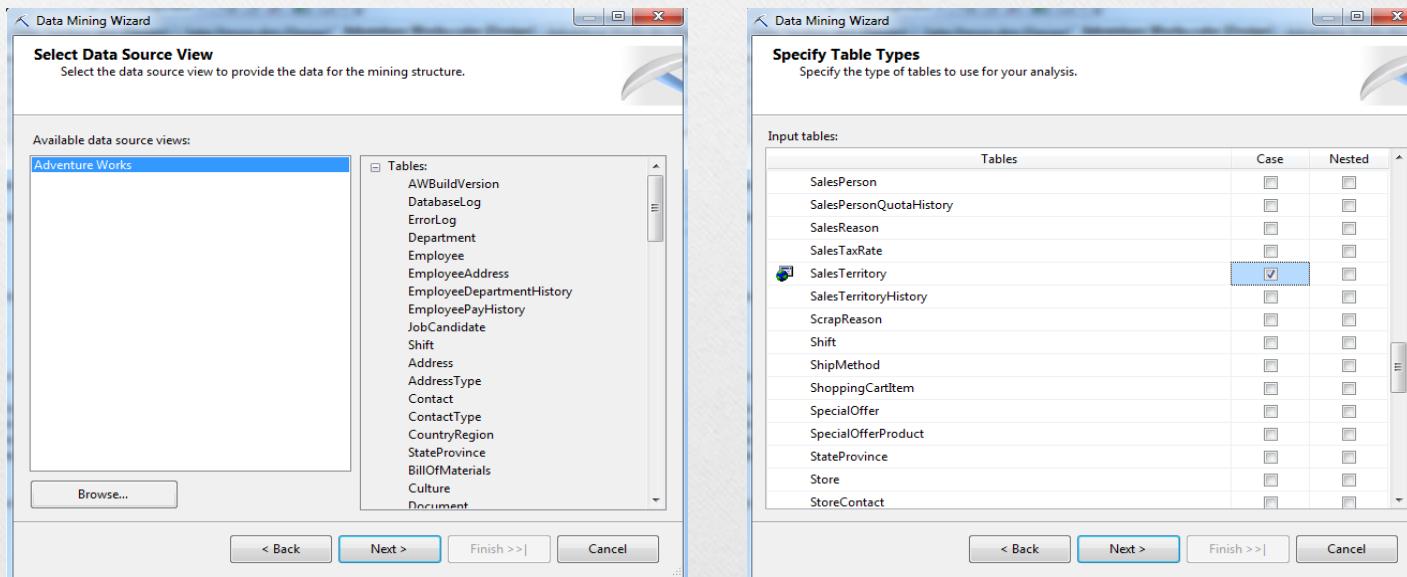
- Data Mining Wizard (cont.)
- create the Data Mining Structure
 - select Microsoft Clustering or Decision Trees algorithm for clustering and classification.



- Microsoft Clustering

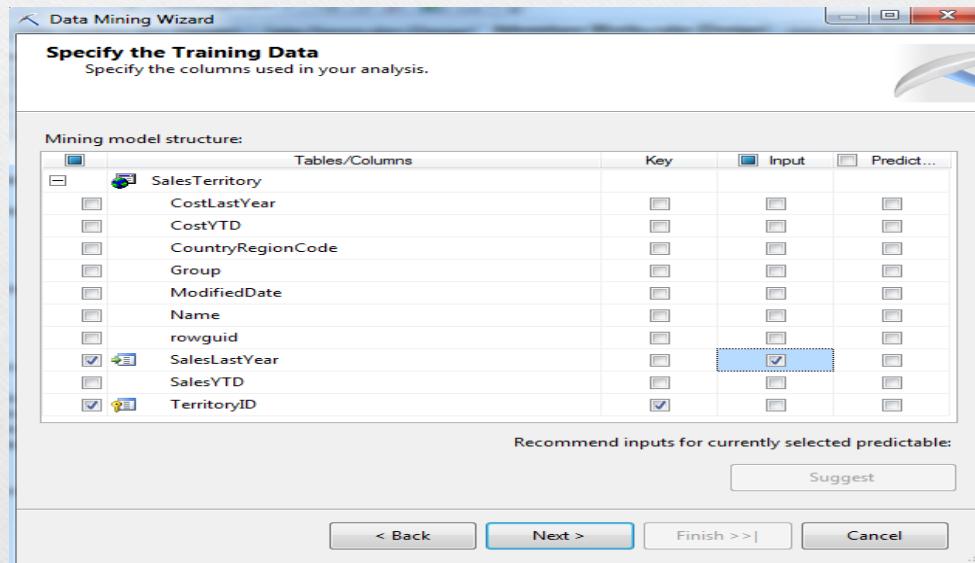
- Select Data Source
 - Adventure Works
- Specify the table type

- select the case table from some defined views,
- whether the table is nested or not.
- In our case, select **SalesTerritory** table for clustering analysis.



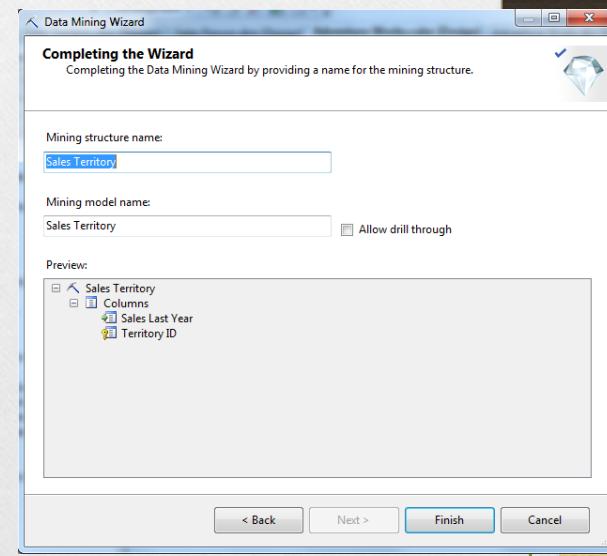
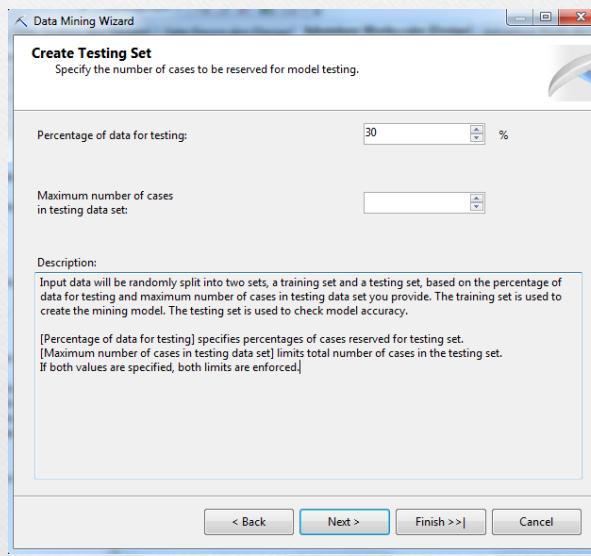
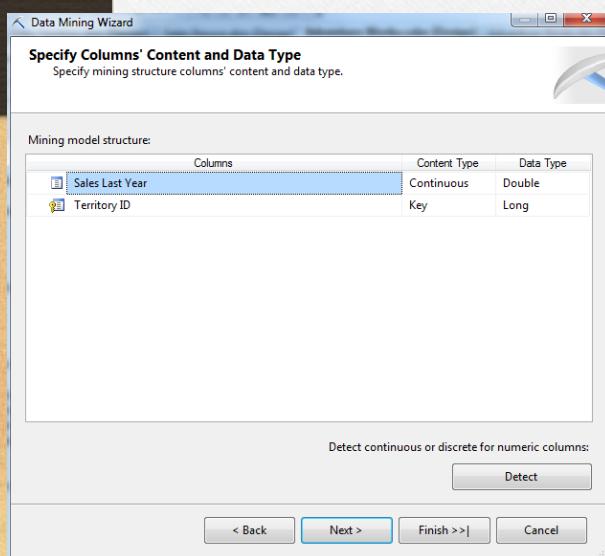
• Microsoft Clustering (cont.)

- Specify the Training data
 - specify the key, input or predictable one from the attributes.
 - if we would like to **cluster territories based on their sales last year**,
 - **TerritoryID** will be the key
 - **SalesLastYear** is the input for providing information to clustering.
 - Note that in the case of the clustering algorithm, there is no need for a predictable attribute.



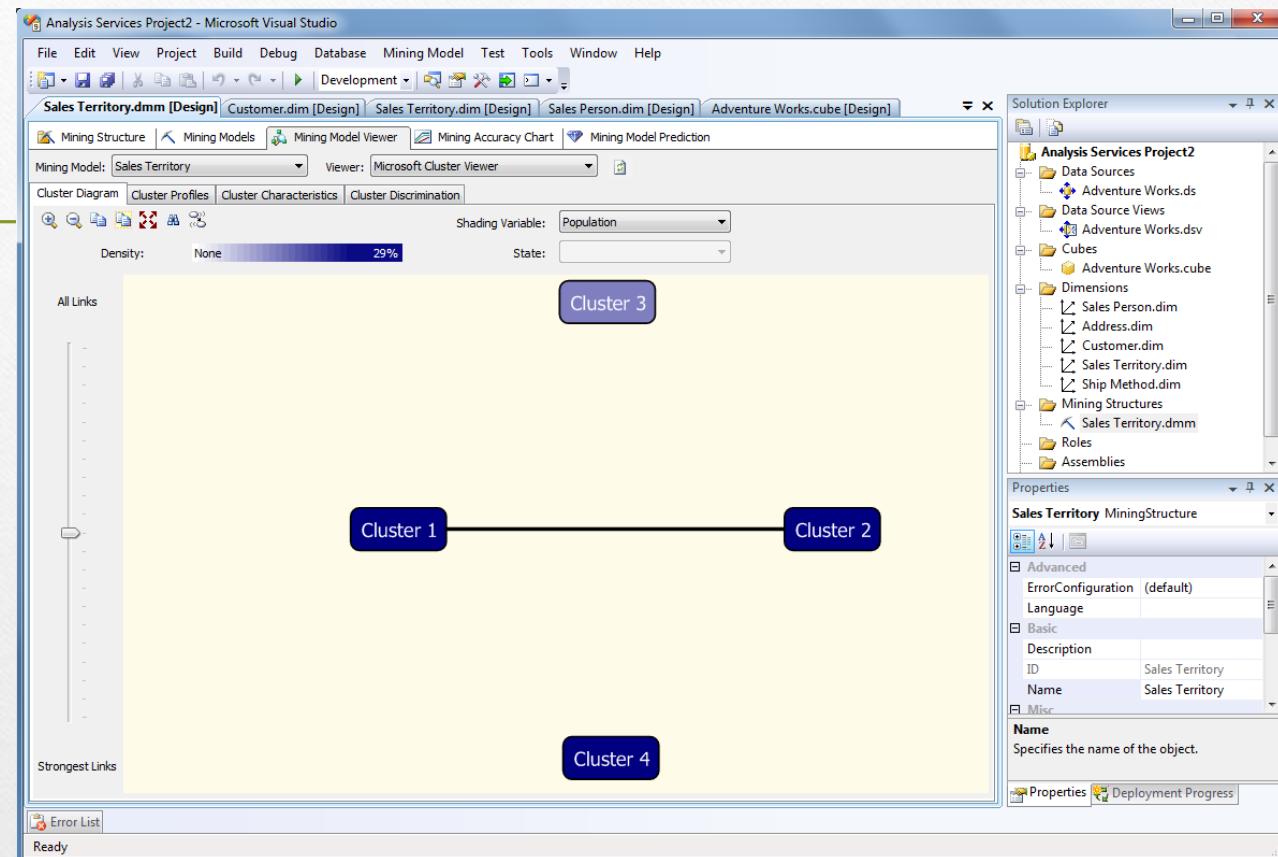
• Microsoft Clustering (cont.)

- specify the column content and data type
- Create testing data set
- Finally, Set a descriptive name for the data mining structure and data mining model container.



DM models in SQL Server

- To process the clustering object, right-click “**Process**”. After clustering object is processed, then we can view the clustering result through the cluster viewer. Right-click clustering object and select “**Browse**”.
- **The cluster viewer**
 - The Microsoft Cluster Viewer provides the following tabs for use in exploring clustering mining models:
 - Cluster Diagram
 - Cluster Profiles
 - Cluster Characteristics
 - Cluster Discrimination



Browsing clustering results

Microsoft Decision Tree

- The steps to create decision tree object are pretty similar to those in creating clustering object outlined above. In the case of classification, we select **SalesOrderHeader** as the table for analysis. **We would like to predict the total sale using two attributes, OnlineOrderFlag and TerritoryID by building a decision tree.**
- In this case, **SaleOrderID** attribute is the key for identifying each case, the **OnlineOrderFlag** and **TerritoryID** are used as the input attributes and **TotalDue** is used as the predictable attribute. Note that in the case of the classification algorithm, you need to specify a predictable attribute.
- And finally the wizard ends after setting a descriptive name for the data mining structure and data mining model container.
- Also, after the classification objects have been created, we need to process these objects to generate decision tree results. Right click the relevant data mining objects and select “**Process**”.

The Tree Viewer

- The Microsoft Tree Viewer for viewing decision tree built includes the following tabs and panes:
 - Decision Tree
 - Dependency Network
 - Mining Legend

Analysis Services Project2 - Microsoft Visual Studio

Sales Order Header 1.dmm [Design] Sales Order Header.dmm [Design] Sales Territory.dmm [Design] Customer.dim [Design]

Mining Structure Mining Models Mining Model Viewer Mining Accuracy Chart Mining Model Prediction

Mining Model: Sales Order Header 1 Viewer: Microsoft Tree Viewer

Decision Tree Dependency Network

Histograms: 6 Background: All Cases Default Expansion: 3 Levels Show Level 1 Level 3

```

graph TD
    All[All] --> OOF[Online Order Flag = False]
    All --> NOOFL[Online Order Flag not = False]
    OOF --> TID9[Territory ID >= 9]
    OOF --> TID6[Territory ID = 6]
    NOOFL --> TID79[Territory ID >= 7 and < 9]
    NOOFL --> TID6L[Territory ID < 6]
  
```

Properties Deployment Solution Explorer

Solution Explorer:

- Analysis Services Project2
 - Data Sources
 - Adventure Works.ds
 - Adventure Works.dsv
 - Cubes
 - Adventure Works.cube
 - Dimensions
 - Sales Person.dim
 - Address.dim
 - Customer.dim
 - Sales Territory.dim
 - Ship Method.dim
 - Mining Structures
 - Sales Territory.dmm
 - Sales Order Header.dmm
 - Sales Order Header 1.dmm
 - Roles

Mining Legend:

High	Low
Total Cases: 22026	
Term	Coeffici...
Territory ID	8128.740
*	-603.056

All
Existing Cases: 22026
Missing Cases: 0

Error List Ready

Thank You
